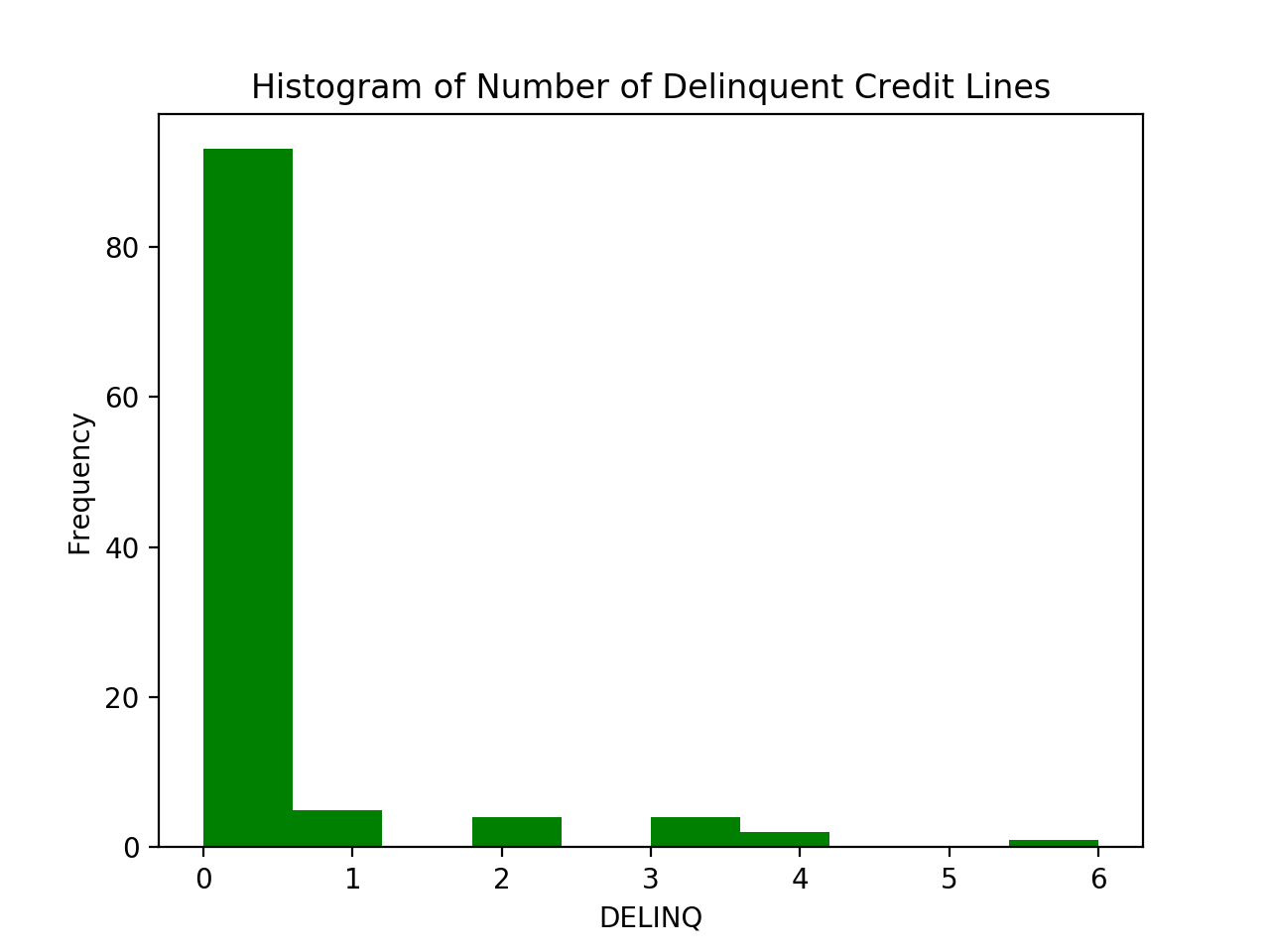
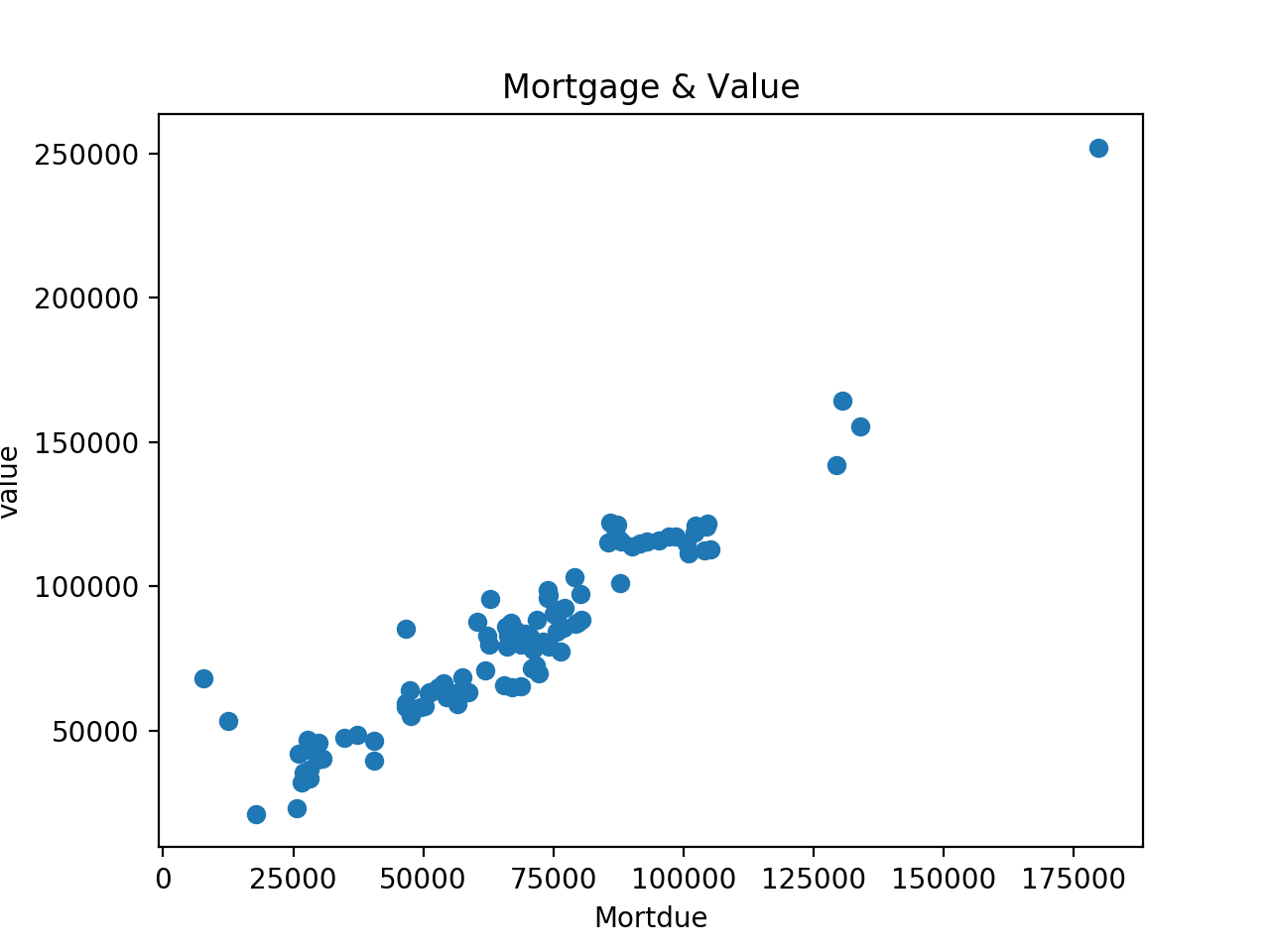
**Shahzeb Khalid**

**CIS 362**

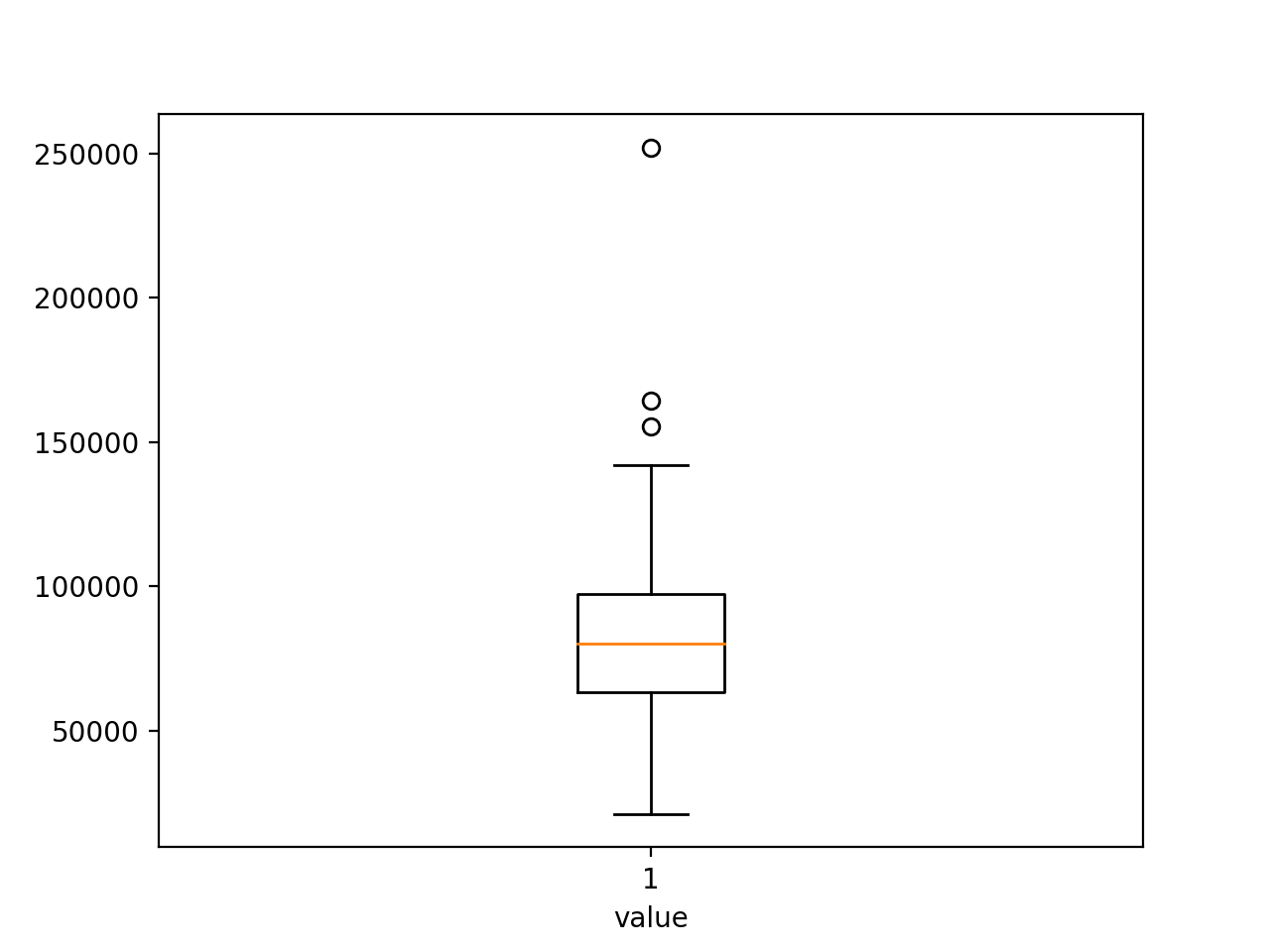
**Descriptive statistics: tabular and graphical representation.**

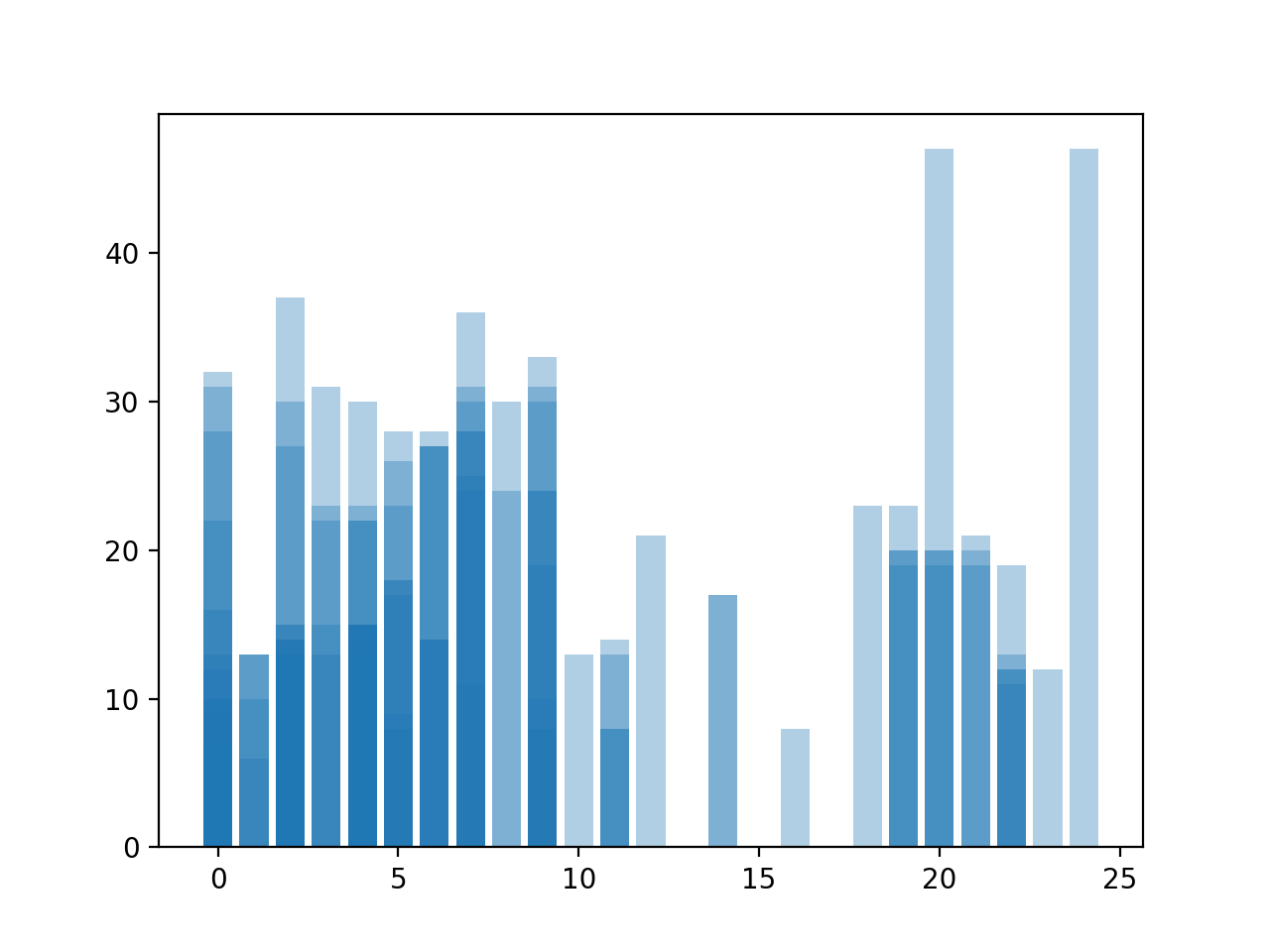
****

**DELINQ:** The distribution is skewed to the right, which also means that it is positive. Also indicating that the mean is greater than the median. In the histogram, it shows us that the number of people whose credit line is really close to zero. This shows us that they have a really good credit line. There a significate amount of people with good credit score lines. However, there are some who are outliers and they are only a small number of people whose credit line is not as impressive compared to those who were in the lower part of the plot. There are few more compared to those where were close to six, there’s credit line is ok.



**Mortgage & value:** I have chosen two variables as correlation, which are mortgage & value. Correlation usually calculates the strength of the relationship between the relative movements of the two variables. When we observe the scatter plot data we have two variables that have a positive correlation for both Mortgage and value. Since the correlation we have is 0.9398 indicating that this is the strongest correlation we have. Also when we observe the plot there’s an outlier. Which lies outside the overall pattern of distribution.

**Value Box Plot:**  : This is used to present the shape of a distribution. This box is best represented for the use of outliers. The line underneath the box is our lower whisker. Which is indicating the lower value of house. As for the box that there are major data lies. Also, the line in the middle of the box that is our median. Also, there are some outliers in the box plot. Which are way up top, they are few of a number of people who’s house values are quite high. Which is really far away from our median or the box. If we were to include the outliers in the upper whisker the line would be really long and the data representation would be off from the most of the data. It would affect the mean value in the data.



**YOJ and CLNO:** The light blue bar is representing the YOJ and the clno is being represented by the dark blue lines. This bar plot shows us how over the years, the number of people credits lines they have. As we can observe that there are not as many people who have many credit line between somewhere 12 to 18. Most of the people fall below 11 and it only decreased over the years. When you notice from 19 to 22 there are significantly fewer people from the 0 – 11, and there where the median also lies as well.

**Frequency Distribution and Percent Frequency Distribution:**

**Bad Frequency**

0 79%

1 21%

**JOBS Frequency**

Office 22%

Manger 6%

Prof 19%

Self 1%

Sales 10%

Other 42%

**REASON** **Frequency**

HomeImp 75%

DebtCon 25%

**Delinq** **Frequency**

* 1. 90%

2-3 7%

4-5 2

5-6 1%

A percent frequency distribution of data that specifies the percentage of observations that exist for each data point. It is most helpful when expressing the relative frequency of the surveys responses. Sometimes they can be represented or displayed as a table or a bar plots or pie charts.